yang_seonhyeHW25

Seonhye Yang

3/31/2019

##Question 1

```
library(faraway)
```

##Question 1

```
attach(pvc, warn.conflicts = F)
#mean and standard deviation for operator
tapply(psize, operator, mean)
```

```
##        1        2        3
## 32.94375 32.68125 31.43750
```

```
tapply(psize, operator, sd)
```

```
##        1        2        3
## 2.722981 2.628363 2.815641
```

```
#mean and standard deviation for resin
tapply(psize, resin, mean)
```

```
##        1        2        3        4        5        6        7        8
## 35.65000 34.61667 29.85000 29.46667 30.85000 30.20000 32.73333 35.46667
```

```
tapply(psize, resin, sd)
```

```
##         1         2         3         4         5         6         7
## 0.6920983 1.0264827 1.5706686 1.0801234 1.2926717 0.8532292 1.7648418
##         8
## 1.8511258
```

##Question 2

```
no_interaction <- lm(psize~resin+operator, data = pvc) #main effect
interaction <- lm(psize~resin*operator, data = pvc) #interaction

summary(no_interaction)
```

```
##
## Call:
## lm(formula = psize ~ resin + operator, data = pvc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9500 -0.6125 -0.0167  0.4063  3.6833
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  36.2396     0.5226  69.345  < 2e-16 ***
## resin2       -1.0333     0.6610  -1.563  0.12630
## resin3       -5.8000     0.6610  -8.774 1.13e-10 ***
## resin4       -6.1833     0.6610  -9.354 2.11e-11 ***
```

```
## resin5       -4.8000     0.6610  -7.261 1.09e-08 ***
## resin6       -5.4500     0.6610  -8.245 5.46e-10 ***
## resin7       -2.9167     0.6610  -4.412 8.16e-05 ***
## resin8       -0.1833     0.6610  -0.277  0.78302
## operator2    -0.2625     0.4048  -0.648  0.52059
## operator3    -1.5063     0.4048  -3.721  0.00064 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.145 on 38 degrees of freedom
## Multiple R-squared:  0.8595, Adjusted R-squared:  0.8262
## F-statistic: 25.82 on 9 and 38 DF,  p-value: 1.474e-13
```

```
summary(interaction)
```

```
##
## Call:
## lm(formula = psize ~ resin * operator, data = pvc)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7000 -0.3625  0.0000  0.3625  2.7000
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       36.2500     0.8598  42.164  < 2e-16 ***
## resin2            -1.1000     1.2159  -0.905 0.374615
## resin3            -5.5500     1.2159  -4.565 0.000126 ***
## resin4            -6.5500     1.2159  -5.387 1.56e-05 ***
## resin5            -4.4000     1.2159  -3.619 0.001372 **
## resin6            -6.0500     1.2159  -4.976 4.42e-05 ***
## resin7            -3.3500     1.2159  -2.755 0.011014 *
## resin8             0.5500     1.2159   0.452 0.655078
## operator2         -0.8500     1.2159  -0.699 0.491216
## operator3         -0.9500     1.2159  -0.781 0.442245
## resin2:operator2   1.0500     1.7195   0.611 0.547175
## resin3:operator2  -0.2000     1.7195  -0.116 0.908372
## resin4:operator2   1.2000     1.7195   0.698 0.491960
## resin5:operator2   0.4000     1.7195   0.233 0.818024
## resin6:operator2   1.3000     1.7195   0.756 0.456985
## resin7:operator2   0.4500     1.7195   0.262 0.795782
## resin8:operator2   0.5000     1.7195   0.291 0.773715
## resin2:operator3  -0.8500     1.7195  -0.494 0.625567
## resin3:operator3  -0.5500     1.7195  -0.320 0.751842
## resin4:operator3  -0.1000     1.7195  -0.058 0.954105
## resin5:operator3  -1.6000     1.7195  -0.931 0.361376
## resin6:operator3   0.5000     1.7195   0.291 0.773715
## resin7:operator3   0.8500     1.7195   0.494 0.625567
## resin8:operator3  -2.7000     1.7195  -1.570 0.129454
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.216 on 24 degrees of freedom
## Multiple R-squared:  0.8999, Adjusted R-squared:  0.804
## F-statistic: 9.382 on 23 and 24 DF,  p-value: 3.447e-07
```

##Question 3

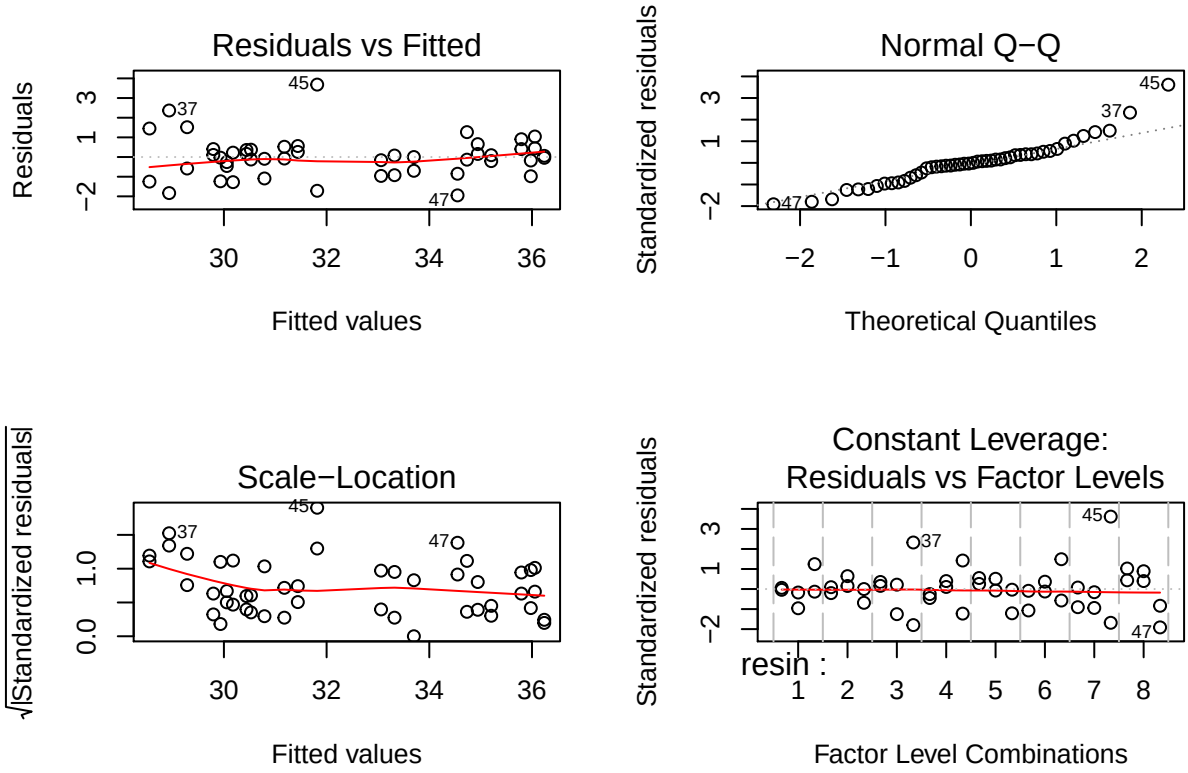
```
anova(interaction)
```

```
## Analysis of Variance Table
##
## Response: psize
##                Df  Sum Sq Mean Sq F value    Pr(>F)
## resin           7 283.946  40.564 27.4388 5.661e-10 ***
## operator        2  20.718  10.359  7.0072   0.00401 **
## resin:operator 14  14.335   1.024  0.6926   0.75987
## Residuals      24  35.480   1.478
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is 0.75987 which is greater than $\alpha = 0.05$ so we cannot reject the null hypothesis (no interaction). Therefore, we can conclude that the it's better to use no interaction model than the interaction model.

##Question 4

```
par(mfrow=c(2, 2))
plot(no_interaction)
```



There seems to be a linear relationship by looking at the Residuals vs Fitted (constant variance is satisfied), there seems to be a normal distribution, the residuals are spread equally along the ranges of predictors which implies that the assumption of equal variance (homoscedasticity) is satisfied and by looking at the Constant Leverage plot, there are no signs of outliers. But just incase, we will perform a bonferroni test to test for outliers.

```
critval = qt(0.05/(2*nobs(no_interaction)), df=df.residual(no_interaction)-1, lower=FALSE)
which(abs(rstudent(no_interaction)) > critval)
```

```
## 45
```

```
## 45
```

However, performing the bonferroni test suggests that 45 is an outlier.

##Question 5

```
CI = TukeyHSD(aov(no_interaction,data=pvc))
CI
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = no_interaction, data = pvc)
##
## $resin
##           diff        lwr        upr     p adj
## 2-1 -1.0333333 -3.1522815  1.0856149 0.7683288
## 3-1 -5.8000000 -7.9189482 -3.6810518 0.0000000
## 4-1 -6.1833333 -8.3022815 -4.0643851 0.0000000
## 5-1 -4.8000000 -6.9189482 -2.6810518 0.0000003
## 6-1 -5.4500000 -7.5689482 -3.3310518 0.0000000
## 7-1 -2.9166667 -5.0356149 -0.7977185 0.0019046
## 8-1 -0.1833333 -2.3022815  1.9356149 0.9999924
## 3-2 -4.7666667 -6.8856149 -2.6477185 0.0000003
## 4-2 -5.1500000 -7.2689482 -3.0310518 0.0000001
## 5-2 -3.7666667 -5.8856149 -1.6477185 0.0000379
## 6-2 -4.4166667 -6.5356149 -2.2977185 0.0000018
## 7-2 -1.8833333 -4.0022815  0.2356149 0.1127668
## 8-2  0.8500000 -1.2689482  2.9689482 0.8984776
## 4-3 -0.3833333 -2.5022815  1.7356149 0.9989372
## 5-3  1.0000000 -1.1189482  3.1189482 0.7958917
## 6-3  0.3500000 -1.7689482  2.4689482 0.9994110
## 7-3  2.8833333  0.7643851  5.0022815 0.0022073
## 8-3  5.6166667  3.4977185  7.7356149 0.0000000
## 5-4  1.3833333 -0.7356149  3.5022815 0.4375901
## 6-4  0.7333333 -1.3856149  2.8522815 0.9507745
## 7-4  3.2666667  1.1477185  5.3856149 0.0003909
## 8-4  6.0000000  3.8810518  8.1189482 0.0000000
## 6-5 -0.6500000 -2.7689482  1.4689482 0.9741405
## 7-5  1.8833333 -0.2356149  4.0022815 0.1127668
## 8-5  4.6166667  2.4977185  6.7356149 0.0000007
## 7-6  2.5333333  0.4143851  4.6522815 0.0098978
## 8-6  5.2666667  3.1477185  7.3856149 0.0000000
## 8-7  2.7333333  0.6143851  4.8522815 0.0042481
##
## $operator
##         diff       lwr        upr     p adj
## 2-1 -0.26250 -1.249747  0.7247472 0.7943575
## 3-1 -1.50625 -2.493497 -0.5190028 0.0018126
## 3-2 -1.24375 -2.230997 -0.2565028 0.0106800
```

```
which(CI$resin[, 4] <= 0.05|CI$resin[, 2] >= 0)
```

```
## 3-1 4-1 5-1 6-1 7-1 3-2 4-2 5-2 6-2 7-3 8-3 7-4 8-4 8-5 7-6 8-6 8-7
##   2   3   4   5   6   8   9  10  11  17  18  21  22  25  26  27  28
```

```r
which(CI$operator[, 4]<= 0.05|CI$operator[, 2] >= 0)
```

```
## 3-1 3-2
##   2   3
```

For resin, 3-1, 4-1, 5-1, 6-1, 7-1, 3-2, 4-2, 5-2, 6-2, 7-3, 8-3, 7-4, 8-4, 8-5, 7-6, 8-6 and 8-7 these are the significant pairs because they have a pvalue less than 0.05 and the CI don't include 0. Also for operator, 3-1 3-2, these are the significant pairs because they have a pvalue less than 0.05 and the CI don't include 0.